DOCUMENT RESUME

ED 093 929 95 TM 003 747

AUTHOR Cooley, William W.
TITLE Assessment of Educational Effects: An Educatiopal Psychologist's Point of View.
INSTITUTION Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE [Apr 74]
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (59th, Chicago, Illinois, April 1974)

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS *Educational Assessment; Educational Programs; Educational Psychology; Educational Research; Evaluation Techniques; Learning Processes; *Program Effectiveness; Program Evaluation
IDENTIFIERS *Educational Outcomes

ABSTRACT

The contribution that educational psycholpgists can make to the improvement of the assessment of educational effects are discussed. Examined are ways in which current psychological knowledge, particularly psychometrics and learning theory, is relevant to: the selection of appropriate criterion measures, the measurement of educational processes, the description of the initial status of the learner, and the analysis of field data. Predictive validity, treatability, and parsimony--three key principles of criteria selection--suggest that the most important criterion for assessing educational programs is general intellectual development. Refered to is work on process measurement in which the author proposes a model of classroom learning which identifies four major process dimensions which assess the opportunity for learning, the degree to which the environment enhances motivation to learn, the quality of the structure of the curriculum, and the effectiveness of the instructional events. To measure the initial status of the learner, the author suggests measuring the dimensions of individual differences that are expected to be affected by the program being assessed, prior to its initiation. Characteristics of an effective statistical model for deriving useful information from these observations are discussed. (Author/RC)

ED 093929
TM 003 247

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Assessment of Educational Effects: An Educational Psychologist's Point of View[1]

William W. Cooley
Learning Research and Development Center
University of Pittsburgh

Since about 1968, there has been a very clear shift in emphasis from federal support of new educational programs to studies concerned with determining the effectiveness of such programs. However, many are now asking whether our present evaluation capability, both in terms of available technique and technicians, can respond adequately to the extraordinary demands being placed upon it.[2] In the present climate, it is extremely appropriate to examine our present ability to attribute particular educational effects to particular educational practices.

---

[1] Invited paper presented as part of "Assessment of Educational Effects: A Multi-Disciplinary View," a symposium held at the meeting of the American Educational Research Association, Chicago, April 1974. This paper was prepared under the auspices of the Learning Research and Development Center, supported in part by funds from the National Institute of Education, United States Department of Health, Education, and Welfare.

[2] For example, the National Advisory Council on Education Professions Development has called for "a full-scale examination of the concepts, methods and manner of conducting evaluations of Federal programs in education [1973, p. 5]."

One way in which some people continue to talk about the quality of educational programs is in terms of school inputs, not effects. That is, schools are compared in terms of cost per pupil, class size, salary schedules, and so on. The assumption is that such variables are clearly related to quality education, and that real effects either cannot or need not be assessed. Recent compilations (e.g., Averch et al., 1972; Jencks et al., 1972) of the research relating these kinds of inputs to achievement outcomes make this assumption appear to be untenable. If you want to know how schools are affecting children, it is necessary to look at outcomes, not budgets.

The National Center for Educational Statistics (1973) recently published a statistical summary of indicators of educational outcomes. Here we find statistical indices that reflect variables with real social significance, such as adult literacy, unemployment, income, mental and physical health, and percent of our population who are imprisoned. Among the 58 such indicators that the Center has summarized, everyone should be able to identify indicators that they consider to be important measures of educational effects. The difficult part is to link such outcomes to particular kinds of educational practices, or even to confidently relate the variance in such indicators to the variance in the amount of education particular individuals received. Although longitudinal studies that follow students from particular educational programs into life outcomes can be revealing, such long-term studies are not likely to solve the immediate policy questions with which effectiveness studies must deal.

In 1966, the Equality of Educational Opportunity study attempted to relate the quality of educational inputs to the quality of student outcomes (Coleman et al., 1966). Although this massive effort represents a landmark in educational research, its shortcomings

are many, including: inadequacy of the measures of both educational process and student outcome, the inherent ambiguity of a cross-sectional rather than longitudinal study, the serious confusions surrounding the appropriateness of particular statistical models, and the conflict in research objective, that is, causal attribution versus a descriptive survey of the distribution of educational resources. These and other shortcomings have been detailed in other publications (including some by the individuals who conducted the study), so they need not be discussed further here (e.g., Mosteller & Moynihan, 1972). The fact that the Coleman effort influenced educational policy in spite of its defects has been extremely important in stimulating inquiry into the assessment of educational effects. The truth of the matter is that assessing educational effects is not nearly so simple as is implied by most evaluation theorists. The recent General Accounting Office (1973) report that is critical of the evaluation activities of federally supported educational labs and centers also gives the impression that evaluation is a rather straightforward task, and why don't we get on with it!

Assessing the effects of educational programs is a research task, and it is no more straightforward than any other research task. To answer evaluation questions with minimum ambiguity requires the same creative talent as the testing of scientific hypotheses. In addition, the evaluation task is confounded by all kinds of practical problems which most researchers can control in laboratory situations but with which the assessor of educational effects must deal in the field. Sanders and Guba (1973) recently proposed a three-dimensional matrix of the kinds of problems the evaluator can encounter. Their cube has 648 cells!

I will not deal here with the practical problems of the educational evaluator, but rather with the contributions that educational psychologists can make to the improvement of our ability to assess educational effects. What follows is an examination of the ways in which current psychological knowledge, particularly psychometrics and learning theory, is relevant to: the selection of appropriate criterion measures, the measurement of educational processes, the description of the initial status of the learner, and the analysis of field data. When I use "we," it is not editorial, but refers to the collaboration with Paul Lohnes that I have profited from over the years, including our current effort to be entitled, Evaluative Inquiry in Education.

What Effects Are Worth Assessing?

If one sets out to assess educational effects, one of the first and most difficult decisions to make concerns what effects to assess. The selection of criteria for educational practices is not susceptible to simple technical solutions. If the assessor is attempting to develop information that is relevant to particular policy decisions, it is critically important that the selection of outcome measures be the result of a dialogue between the evaluator and the consumer of the evaluation results. The situation is complicated in most policy research by the multiplicity of consumers with varying values.

For outcomes to be valued, they must be perceived as a link in a means-ends continuum leading to a desired end-in-view. The most important school outcomes will generally be those that people believe affect, either directly or indirectly, success and satisfaction as an adult. Longitudinal studies such as Project TALENT have begun to reveal the importance of certain predictors of career development. Evidence of this kind helps to establish the link between

attributes developed in school and post-school performance. Predictors with established validity for determining post-school adjustments can be justified as criteria for assessing program effects without the need for long-term longitudinal studies for each policy question.

There are two other considerations involved in the selection of school outcome measures. First of all, there must be some theoretical or empirical basis for expecting that the outcome measures can be affected by the educational practices being assessed. Secondly, redundancies among outcome measures must be reduced to a minimum for ease in describing and interpreting the results.

These three key principles of criteria selection--predictive validity, treatability, and parsimony--suggest that the most important criterion for assessing educational programs is general intellectual development. Here are a few of the reasons why we believe this general factor is so very important.

1. Any battery of cognitive tests given to a sample of subjects from a heterogeneous population results in a set of positively correlated scores, the principal component of which generally accounts for at least one-fourth of the variance in the original measurements. This principal component is a general measure of an individual's current profile level on that set of tests.

2. If one administers two different batteries of cognitive tests, the principal component from one battery will generally correlate at least .8 with the principal component from the other battery.

3. The general factor, when measured at one educational level, is by far the best predictor of academic performance at the

next level. What individuals are able to learn today is mainly a function of what they have learned to date.

4. General intellectual development is, in part, a function of school practices. Although half its variance may be attributable to prior intellectual development, half is not. We are now finding ways of attributing some of the variance unexplained by prior development to different educational practices.

5. The general factor is by far the best predictor of what happens to youth upon leaving school. It is, for example, the best single predictor of the quality of the vocational prizes that one achieves.

Although general intellectual development can be justified as the primary measure of school effects, it is clearly not the only factor that is of interest. However, to keep the number of criteria to a manageable size, to eliminate redundancies among criteria, and to reduce interpretive ambiguities that result from highly correlated outcome measures, we emphasize the utility of an orthogonal set of general factors. Such a set of uncorrelated dimensions preserves most of the currently measurable variance in student differences, is based upon decades of psychometric research, and has known or knowable predictive validities.

Elsewhere (Cooley & Lohnes, in preparation), we have drawn upon our Project TALENT research (Lohnes, 1966; Cooley & Lohnes, 1968) to illustrate such a multidimensional representation of educational outcomes that satisfies these conditions. Although trait and factor theory today is not mainstream psychological thought, we do believe that it will prevail as the basis for solving the very practical problem of representing educational outcomes, just as it has prevailed in the solution of other practical problems. The most critical need in improving our ability to assess educational effects is to

develop a more adequate basis for representing a broad spectrum of student outcomes, and to demonstrate their extra-school transfer value.

You will note that we have emphasized general measures of educational outcomes. Such measures have been criticized because they lack diagnostic value for the individual student. Knowing where Johnny is on a principal component of general intellectual development, for example, is not too useful in planning his current work in mathematics. In fact, knowing where he is on a general mathematics factor is not even too useful for that purpose. The general measures are important, however, because they are the kinds of measures for which one can establish extra-school predictive validities. Combined with research on instructional processes, these measures can have excellent diagnostic value for educational programs.

What Educational Practices Should Be Assessed?

One principle of evaluative research that has become extremely clear in recent years is that the educational processes being assessed cannot be expected to be implemented uniformly across students, classrooms, schools, etc. In fact, variation in the implementation of a given program can be so great that its overlap with a competing program may make it meaningless to contrast the effects of the two programs. Not only is there variation in how an innovation is implemented, but as Charters and Jones (1973) recently pointed out so well, the innovation may even be a "non-event." For these and other reasons, the actual educational process under investigation must be directly observed and then represented as a multidimensional domain in the same way that one must consider outcomes. Although doing so implies an additional expense for the assessment of

educational effects, studies that ignore this variation in implementation are going to be seriously ambiguous.

Another advantage of directly observing school process is that it allows one to move from assessment of the effects of specific educational programs, from which we learn very little, to assessment of more general educational practices, from which we can learn a great deal. Classroom practices, measured in terms of dimensions derived from a theory of instruction, are likely to be more important than differences among specific educational programs. For example, the current review article, "Comparing Curricula," reminds us that students will tend to learn that which is included in their coursework better than that which is not (Walker & Schaffarzick, 1974)! So contrasting curricula effects, even if there are differences, tells us little more than was already known, namely the content differences of the various curricula. Still another argument for good process description is that since value is attached to educational means as well as ends, process information is just as important as the relation between process and outcome.

To guide this work on process measurement, we (Cooley & Lohnes, in preparation) propose a model of classroom learning which identifies four major process dimensions derived from Carroll's (1963) model. Briefly summarized, they assess the opportunity for learning, the degree to which the environment enhances motivation to learn, the quality of the structure of the curriculum, and the effectiveness of the instructional events. Assessing the learning environment in this way, and combining these four dimensions with the abilities and motives with which a student enters the educational experience being assessed, will explain most of the variance in educational outcomes.

What Needs to Be Known About the Learner?

One of the best established, yet frequently ignored principles in the assessment of educational effects is that the state of the students' abilities and motives as they enter an educational program is always the strongest predictor of what they will achieve in that program. The most obvious way to deal with this problem is to measure the dimensions of individual differences that are expected to be affected by the program being assessed, prior to the initiation of that program. This, of course, is not exactly a novel idea! But of the nineteen effectiveness studies recently summarized by Averch et al. (1972), only one actually did this. To ignore measurement of the initial status of the learner results in the same kind of ambiguity as ignoring the measurement of process or the measurement of outcomes.

The need for measuring the initial status of the learner, the process dimensions of the learning environments, and the status of the learner at the end of the process being assessed, is why we encourage the conduct of short-term longitudinal studies for the assessment of educational effects. Although random assignment to treatment is not a necessary aspect of this approach, measurement of all three domains in a longitudinal fashion is essential. The known transfer value of the outcome measures is what makes it possible for studies to be short term. However, educational programs worthy of such assessment, and the general nature of the important criteria, suggest studies of at least one year's duration.

How Should the Data Be Analyzed?

Given the three multidimensional domains that summarize the variance which occurs in the initial status of the learner, the educational processes to which the learner is exposed, and the learning

outcomes, a remaining problem is the choice of an appropriate statistical model for deriving useful information from these observations. What the assessor of educational effects generally must settle for is a research design that is less controlled than the laboratory experiment, but that need not be so chaotic as is implied by the notion of "nature's experiment." What we recommend is the method of controlled correlation, a type of quasi-experiment. Generally, some degree of control is possible over what happens in schools and classrooms, and this control can be taken advantage of in order to reduce the correlations among independent variables, such as result in a purely naturalistic field study.

A statistical model that is capable of sorting out the relative impact of initial status and process dimensions on outcome is a correlation/regression approach that includes partitioning of the variance explained into unique and common contributions for the initial status and process dimensions. This commonality technique was popularized by Mayeske et al. (1969) in their re-analysis of the Coleman data. Instead of saying, as Coleman did, that school practices do not seem to have much effect on outcome, their re-analysis shows that families and schools are so assortatively mated in American society that most of their influence on academic achievement is inseparable in uncontrolled survey data. This high correlation between the socioeconomic status of the learner and effective educational practices, which makes it impossible to sort out school effects from home effects, can be reduced by planned intervention. This planned variation in treatment can be applied iteratively from one school year to the next, with interventions modifying process in order to reduce the correlations among the process dimensions and between process and initial status of the learner. By manipulating process in this way, and by keeping the correlations

among the independent variables low, we can achieve what is implied by the assessment of effects, which is causal attribution. In the laboratory, all this is achieved by random assignment of subjects to treatments in an orthogonal design. But in the field, there is great resistance to randomly assigning children to different educational environments (e.g., schools, classrooms, peer groups, families!), and with implementation variation, the design does not stay orthogonal anyway.

Have These Notions Been Applied?

What I have outlined here are some considerations relevant to the assessment of educational effects. I have not presented a full-blown evaluation model. My impression is that our literature already has an adequate supply of evaluation models. Thus, it didn't seem useful to add to that abundance. What is not abundant, however, are convincing results. The reason for the lack of results is that we have tended to talk around the heart of the evaluation problem, which is to conduct studies and analyze data so that particular educational effects can be attributed to particular educational practices. Until we do this well, most evaluation activity will be an empty exercise.

Fortunately, I can point to two efforts that illustrate the approach I've just outlined. In the first of these, Leinhardt (1974) studied the process variation occurring in 52 second-grade classrooms, all of which were implementing a program of adaptive education developed at the University of Pittsburgh's Learning Research and Development Center. Organizing her process measures into four sets of variables suggested by our modification of Carroll's (1963) model of classroom instruction, she found that the process variance within the instructional program uniquely explained about 14 percent of the variance in end-of-year school achievement in the presence of the initial abilities of the children. Some

of her results are clear validations of the components of the instructional program. For example, in classrooms where there was a greater degree of conformity to the developer's instructional model, there was a greater degree of achievement. In a few cases, the results indicated that the instructional model requires modification, in that departures from the model actually enhanced achievement. Leinhardt's field research is essentially an attempt to validate the components of a particular instructional model, and since it generates information that is useful to both the developer and the potential consumer, it is one kind of evaluative inquiry.

A second example of this approach (Cooley & Emrick, 1974) involved re-analyses of the national Follow Through data being collected at the Stanford Research Institute. One can find modest effects for the differences among the programmatic packages developed by different Follow Through sponsors. However, by focusing on dimensions along which classrooms differ, regardless of sponsor, and deriving these dimensions from a model of classroom learning, it is possible to attribute one-fourth of the variance in classroom achievement to variation in classroom processes. This may not be traditional evaluation because it is not directly addressed to policy questions such as, "Should Follow Through be continued?" or "Was program A better than program B?" However, the approach does reveal the effects of different educational practices, such as the fact that the more structured programs are more effective in developing basic academic skills in young children.

These results contrast rather dramatically with studies of schooling effects in which the process measures were based upon information easily available from the principal's office, but had little to do with what was going on in the classroom. The assessment of

educational effects, if it is to provide a basis for improving the quality of the learning experiences of students in schools, must include studies of the type outlined here. They illustrate the best way that this educational psychologist can suggest for attributing particular effects to particular causes. If more creative talent can be brought to bear on this central evaluation task, then educational psychologists will be able to provide useful information for educational policy-makers.

References

Averch, H. A., et al. How effective is schooling? A critical review and synthesis of research findings. Santa Monica, California: Rand Corporation, 1972.

Carroll, J. B. A model of school learning. Teachers College Record, 1963, 64, 723-733.

Charters, W. W., Jr., & Jones, J. E. On the risk of appraising non-events in program evaluation. Educational Researcher, 1973, 2(11), 5-7.

Coleman, J. S., et al. Equality of educational opportunity. Washington, D. C.: United States Government Printing Office, 1966.

Cooley, W. W., & Emrick, J. A. A model of classroom differences which explains variation in classroom achievement. Paper presented at the meeting of the American Educational Research Association, Chicago, April 1974.

Cooley, W. W., & Lohnes, P. R. Predicting development of young adults. Palo Alto, California: Project TALENT, 1968.

Cooley, W. W., & Lohnes, P. R. Evaluative inquiry in education, in preparation.

Jencks, C., et al. Inequality: A reassessment of the effect of family and schooling in America. New York: Basic Books, 1972.

Leinhardt, G. Evaluation of the implementation of a program of adaptive education at the second grade (1972-73). Paper presented at the meeting of the American Educational Research Association, Chicago, April 1974.

Lohnes, P. R. Measuring adolescent personality. Pittsburgh: Project TALENT, 1966.

Mayeske, G. W., et al. A study of our nation's schools. Washington, D. C.: United States O of Educati 969.

Mosteller, F., & Moynihan, D. P. (Eds.) On equality of educational opportunity. New York: Random House, 1972.

National Advisory Council on Education Professions Development. Evaluation of education: In need of examination. Washington, D. C.: Author, 1973.

National Center for Educational Statistics. Indicators of educational outcome, Fall 1972. Washington, D. C.: United States Government Printing Office, 1973.

Sanders, J. A., & Guba, E. G. A taxonomy of problems confronting the practitioner of educational evaluation and a call for the establishment of a national clearinghouse for such problems and potential solutions. Unpublished manuscript, Northwest Regional Laboratory and Indiana University, 1973.

United States General Accounting Office. Educational laboratory and research and development center programs need to be strengthened. Washington, D. C.: GAO, 1973.

Walker, D. F., & Schaffarzick, J. Comparing curricula. Review of Educational Research, 1974, 44(1), 83-111.